Human Mutation

High-throughput Sequencing of a 4.1 Mb Linkage Interval Reveals *FLVCR2* Deletions and Mutations in Lethal Cerebral Vasculopathy

Journal:	*Human Mutation*
Manuscript ID:	humu-2010-0234.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	18-Jun-2010
Complete List of Authors:	THOMAS, Sophie; INSERM U781, Genetics; Université René Descartes ENCHA-RAZAVI, Ferechte; INSERM U781, Genetics; Université René Descartes DEVISME, Louise; CHRU de Lille, Pôle de Pathologie, Centre de Biologie Pathologie Etchevers, Heather; INSERM U-781, Genetics; Université René Descartes Bessieres-Grattagliano, Bettina; ADHMI Institut de Puériculture et de Périnatalogie, Anatomo-foeto-pathologie GOUDEFROYE, Geraldine; Hôpital Necker-Enfants Malades, APHP, Genetics ELKHARTOUFI, Nadia; Hôpital Necker-Enfants Malades, APHP, Genetics Pateau, Emilie; CEA, DSV, Institut de Génomique, Genoscope, Centre National de Séquençage ICHKOU, Amale; Hôpital Necker-Enfants Malades, APHP, Genetics Bonniere, Maryse; APHP Necker, Embryologie; Laboratoire Nord Pathologie Marcorelles, Pascale; CHRU Hôpital Morvan, Anatomopathologie Pathologique PARENT, Philippe; CHRU Hôpital Morvan, Département de pédiatrie et génétique médicale Manouvrier, Sylvie; CHRU de Lille, Hôpital Jeanne de Flandre, Service de Génétique Clinique et Université Lille 2 HOLDER, Muriel; CHRU de Lille, Hôpital Jeanne de Flandre, Service de Génétique Clinique et Université Lille 2 Laquerriere, Annie; CHU Rouen, Anatomie pathologique Loeuillet, Laurence; CHI Poissy, Service d'Anatomie et de Cytologie Pathologiques ROUME, Joelle; CHI Poissy, Génétique Médicale MARTINOVIC, Jéléna; Assistance Publique - Hôpitaux de Paris, APHP, Genetics; Hôpital Necker-Enfants Malades, APHP, Genetics

	MOUGOU-ZERELLI, Soumaya; Université René Descartes; INSERM U781, Genetics; Hôpital Farhat Hached, Génétique moléculaire et Biologie de la reproduction Gonzales, Marie; APHP, Armand Trousseau, Université Pierre et Marie Curie, Paris 6, Génétique et Embryologie Médicales MEYER, Vencent; CEA, DSV, Institut de Génomique, Genoscope, Centre National de Séquençage Wessner, Marc; CEA, DSV, Institut de Génomique, Genoscope, Centre National de Séquençage Bole Feysot, Christine; Fondation IMAGINE, Plateforme de génomique, Nitschke, Patrick; Université Paris - Descartes, Service de Bioinformatique Leticee, Nadia; Hopital Necker-Enfants Malades, APHP, Service de Gynécologie Obstétrique Munnich, Arnold; Hôpital Necker-Enfants Malades, APHP, Genetics; Université René Descartes; INSERM U781, Genetics LYONNET, Stanislas; Hôpital Necker-Enfants Malades, APHP, Genetics; Université René Descartes; INSERM U781, Genetics Wookey, Peter; University of Melbourne, Department of Medicine Gyapay, Gabor; CEA, DSV, Institut de Génomique, Genoscope, Centre National de Séquençage Folliguet, Bernard; Maternite de Nancy, Laboratoire de Biologie de la Reproduction et du Développement VEKEMANS, Michel; Hôpital Necker-Enfants Malades, APHP, Genetics; Université René Descartes; INSERM U781, Genetics ATTIE-BITACH, Tania; INSERM U-781, Genetics; Hôpital Necker-Enfants Malades, Département de Génétique; Université René Descartes
Key Words:	Fowler syndrome, cerebral proliferative vasculopathy, FLVCR2, Hydranencephaly, Fetal lethality, Arthrogryposis

High-throughput Sequencing of a 4.1 Mb Linkage Interval Reveals *FLVCR2* Deletions and Mutations in Lethal Cerebral Vasculopathy

Sophie Thomas[1,2], Ferechté Encha-Razavi[1,2,3], Louise Devisme[4], Heather Etchevers[1,2], Bettina Bessieres-Grattagliano[5], Géraldine Goudefroye[3], Nadia Elkhartoufi[3], Emilie Pateau[6], Amale Ichkou[3], Maryse Bonnière[3,7], Pascale Marcorelle[8], Philippe Parent[9], Sylvie Manouvrier[10], Muriel Holder[10], Annie Laquerrière[11], Laurence Loeuillet[12], Joelle Roume[13], Jelena Martinovic[3], Soumaya Mougou-Zerelli[1,2,14], Marie Gonzales[15], Vincent Meyer[6], Marc Wessner[6], Christine Bole Feysot[16], Patrick Nitschke[17], Nadia Leticee[18], Arnold Munnich[1,2,4], Stanislas Lyonnet[1,2,4], Peter Wookey[19], Gabor Gyapay[6], Bernard Foliguet[20], Michel Vekemans[1,2,4], Tania Attié-Bitach[1,2,4]

1- INSERM U-781, Hôpital Necker-Enfants Malades, Paris, France
2- Université René Descartes, Paris 5, France
3- Département de Génétique, Hôpital Necker-Enfants Malades, AP-HP, Paris, France
4- Pôle de Pathologie, Centre de Biologie Pathologie, CHRU Lille, France
5- Laboratoire d'Anatomo-Foeto-Pathologie, Institut de Puériculture et de Périnatalogie, Paris, France
6- CEA, DSV, Institut de Génomique, Genoscope, Evry, France
7- Laboratoire Nord Pathologie, Lille, France
8 - Laboratoire d'anatomopathologie, CHRU Hôpital Morvan, Brest, France
9 - Département de pédiatrie et génétique médicale, CHRU Hôpital Morvan, Brest, France
10- Service de Génétique Clinique et Université Lille 2, CHRU de Lille, Hôpital Jeanne de Flandre, Lille, France
11- Laboratoire d'Anatomie Pathologique, Hôpital de Rouen, Rouen, France
12- Service d'Anatomie et de Cytologie Pathologiques,CHI Poissy, Saint Germain en Laye, France
13 - Génétique Médicale, CHI Poissy, Saint Germain en Laye, France
14 - Service de Cytogénétique, Génétique moléculaire et Biologie de la reproduction, Hôpital Farhat Hached, Sousse, Tunisie

15- Service de Génétique et d'Embryologie Médicales, Hôpital Armand Trousseau, AP-HP, et Université Pierre et Marie Curie, Paris 6, France

16- Plateforme de génomique, Fondation IMAGINE

17- Service de Bioinformatique, Université Paris - Descartes

18- Service de Gynécologie Obstétrique, Hôpital Necker Enfants Malades, Paris, France

19- Department of Medicine, University of Melbourne, Australia

20- Laboratoire de Biologie de la Reproduction et du Développement, Maternité de Nancy, France

Corresponding author:

Tania ATTIE-BITACH

Département de Génétique et Unité INSERM U-781

Hôpital Necker-Enfants Malades, Paris, France

Tel: 33 (0) 1 44 49 51 44

Fax: 33 (0) 1 44 49 51 50

Email: tania.attie@inserm.fr

ABSTRACT:

Rare lethal disease gene identification remains a challenging issue, but it is amenable to new techniques in high-throughput sequencing (HTS). Cerebral proliferative glomeruloid vasculopathy (PGV), or Fowler syndrome, is a severe autosomal recessive disorder of brain angiogenesis, resulting in abnormally thickened and aberrant perforating vessels leading to hydranencephaly. In 3 multiplex consanguineous families, genome-wide SNP analysis identified a locus of 14 Mb on chromosome 14. In addition, 280 consecutive SNPs were identical in two Turkish families unknown to be related, suggesting a founder mutation reducing the interval to 4,1Mb. To identify the causative gene, we then specifically enriched for this region with sequence capture and performed HTS in a proband of 7 families. Due to technical constraints related to the disease, the average coverage was only 7X. Nonetheless, iterative bioinformatic analyses of the sequence data identified mutations and a large deletion in the *FLVCR2* gene, encoding a twelve transmembrane domain-containing putative transporter. A striking absence of alpha-smooth muscle actin immunostaining in abnormal vessels in fetal PGV brains, suggests a deficit in pericytes, cells essential for capillary stabilisation and remodelling during brain angiogenesis. This is the first lethal disease-causing gene to be identified by comprehensive HTS of an entire linkage interval.

Key words **Fowler syndrome, cerebral proliferative vasculopathy, *FLVCR2*, hydranencephaly, fetal lethality, arthrogryposis**

INTRODUCTION

Cerebral proliferative glomeruloid vasculopathy (PGV) is a severe autosomal recessive disorder of brain angiogenesis, resulting in abnormally thickened and aberrant perforating vessels, forming glomeruloids with inclusion-bearing endothelial cells. This peculiar vascular malformation was delineated by Fowler in 1972 in relation to a stereotyped, lethal fetal phenotype (OMIM 225790), associating hydranencephaly and hydrocephaly with limb deformities (Fowler, et al., 1972). PGV disrupts the developing central nervous system (CNS) but the reason for which abnormal angiogenesis is restricted to the CNS parenchyme remains unknown. Arthryogryposis, when present, appears to be a secondary result of CNS motoneuron degeneration, itself one potential outcome of perfusion failure. Since its earliest description, 42 PGV cases from 26 families have been reported on the basis of histological criteria (Bessieres-Grattagliano, et al., 2009; Williams, et al. 2010).

Identification of a causative gene for a very rare lethal syndrome is a challenge at many levels. The first issue is to find a family that allows the identification of a linkage interval. Such an interval may contain too many genes to make the classical subsequent strategy practical, consisting in designing primers that will permit sequencing of each exon of all the genes of the region. The second difficulty is that sequencing of all the exons is sometimes vain in light of the growing number of non-coding regions identified as pathogenic alleles (Benko, et al., 2009; Kleinjan and van Heyningen, 2005; Lettice, et al., 2003). Finally, for prenatally lethal syndromes such as PGV, technical constraints such as poor quality genomic DNA samples are added. Recent advances in biotechnology permit the sequencing of all the DNA, including the non-coding regions, in most genomic

intervals. After homozygosity mapping of a 4.1 Mb region, we applied targeted genome capture by using a NimbleGen array and high-throughput Roche 454 GS FLX sequencing to the genomic DNA of the proband of 6 families. Bioinformatic analysis of the data allowed us to identify *FLVCR2* (MIM 610865) as the gene responsible for Fowler syndrome (FS). High-throughput sequencing (HTS) generated false positive and false negative results, in part due to insufficient sequencing coverage, and unless care is taken, these can engender the risk of missing mutations during the analysis.

MATERIALS AND METHODS

Patients : The 7 families analysed have been previously reported (Families I to VII (Bessieres-Grattagliano, et al., 2009)). Genomic DNA was extracted from frozen tissue or cultured amniocyte cells in fetal cases and from peripheral blood samples for parents and unaffected siblings.

Genome linkage screening and linkage analysis: Genome-wide homozygosity mapping was performed using 250K Affymetrix SNP arrays in 5 affected and 3 unaffected individuals of two Turkish and one French multiplex, consanguineous families. Data were evaluated by calculating multipoint lod scores across the whole genome using MERLIN software, assuming recessive inheritance with complete penetrance.

NimbleGen Sequence capture and high-throughput sequencing: A custom sequence capture array was designed and manufactured by Roche NimbleGen (Madison, WI, USA). 21 micrograms of genomic DNA was used for sequence-capture in accordance

with the manufacturer's instructions (Roche NimbleGen) and a final amount of 3 micrograms of amplified enriched DNA was used as input for generating a ssDNA library for HTS; 25% lane of a Roche 454 GS FLX sequencer with Titanium reagents) yielding 135 Mb of sequence data per sample.

Capillary sequencing of *FLVCR2* : Primers were designed in introns flanking the 10 exons using the "Primer 3" program (http://fokker.wi.mit.edu/primer3/input.htm) and are listed in Supp. Table S2. PCR were all performed in the same conditions, with a touchdown protocol consisting of denaturation for 30s at 96°C, annealing for 30s at a temperature ranging from 64°C to 50°C (decreasing 1° during 14 cycles, then 20 cycles at 50°) and extension at 72°C for 30s. PCR products were treated with Exo-SAP IT (AP Biotech), and both strands were sequenced with the appropriate primer and the "BigDye" terminator cycle sequencing kit (Applied Biosystems Inc.) and analyzed on ABI3130 automated sequencers. Mutation numbering is based on cDNA reference sequence NM_017791.

Immunohistochemistry : Immunohistochemistry was carried out on six micrometer selected sections using antisera directed against smooth muscle actin (diluted 1:800). Immunohistochemical procedures included a classical microwave pre-treatment protocol in citrate buffer to aid antigen retrieval. Incubations were performed for one hour at room temperature, using the TECHMATE system (DAKOPATTS-Trappes-France). After incubation, histological slides were processed using the LSAB detection kit (DAKOPPATTS-Trappes-France). Peroxidase was visualized by means of either 3-3' diaminobenzidine or amino-ethyl carbazole.

RESULTS

We have collected DNA from fetuses of 7 families reported earlier (Families I to VII (Bessieres-Grattagliano, et al., 2009)). All 14 fetal cases bore the brain-specific angiogenic anomalies characteristic of PGV, resulting in thickened and aberrant perforating vessels and glomeruloids, as exemplified in Figure 1A. Endothelial cells (ECs) were positive for CD34 in both control fetal brains (Figure 1B) and in the tortuous glomerular capillaries (Figure 1C). VEGF-A, while not normally expressed by small brain capillaries (Figure 1D), was strikingly found in the glomerular ECs of PGV fetuses (Figure 1E arrowhead). Like normal Ecs though, PGV ECs expressed VEGFR2 and, weakly, Glut-1 (not shown). CD68, characteristic of macrophages, was completely absent (data not shown). Numerous GFAP-positive astrocytes were observed throughout the cerebral mantle, with immunoreactive endfeet justaposed to glomeruloids (Figure 1F). An antibody to alpha-smooth muscle actin (aSMA) stained vessels within the outer leptomeninges and the walls of perforating vessels in normal fetal brains (Figure 1G). In contrast, although PGV meningeal vessels had similar aSMA expression, the dysplastic intraparenchymous vessels were irregularly stained, if at all (Figure 1H), while most glomeruloid vessels were negative for aSMA (Figure 1I).

To find the molecular basis for this phenotype, we first undertook a genome-wide SNP analysis using an Affymetrix 250K SNP chip with 5 affected and 3 unaffected members of two Turkish and one French multiplex, consanguineous families. Informed consent was obtained from all patients and their relatives; clinical data of all families have previously been reported (Bessieres-Grattagliano, et al., 2009). Genome-wide linkage

analysis conducted with the MERLIN program revealed a 13 Mb genomic region on chromosome 14 from rs10151019 to rs12897284, with a lodscore of 5.4. Moreover, four affected sibs from the two Turkish families shared the same alleles for 280 consecutive SNPs, suggesting a founder effect and reducing the interval to 4.1 MB, from rs2803958 to rs11159220. These two families originated from villages 12 km apart in Khramanmaraps (central Turkey). Microsatellite marker analysis further confirmed the same disease allele in both families, and showed linkage in 3 additional families (Figure 2).

To identify the causative gene, we applied array-based sequence capture of the complete 4.1 MB region followed by high-throughput sequencing. DNA from one proband of 6 families, the heterozygous mother from family I, and a healthy brother not carrying the at-risk allele were selected (Figure 2). Coverage varied from 2X to 12X in individuals depending on the integrity of their DNA (Table 1), with an average coverage depth of 7X; 60% (851,147) of the enriched reads were located on the targeted regions. Only 25% of the targeted regions reached 10X coverage depth.

The number of the detected variations was too large to handle them manually. To facilitate the analysis of these variations a specific genome browser was set up to visualise the locations of variations on the genome, and at the same time an analysis tool has been developed. This analysis tool applied a series of filters to the identified variations. These filters were based on the following criteria: 1) the quality of the sequence variant measured as the number of reads that detected the variant 2) the presence or absence of variants in public databases such as dbSNP and HapMap. 3) the

presence or absence of the variants among the studied samples 4) annotation of the sequence variants based on their location (intron, exon, etc ...) and the characteristics of the resulting change such as synonymous, non-synonymous or stop mutation. Filtered results were visualised in an interactive table permitting us to sort and analyse the results. Thus, initial analysis of the sequence data that met an arbitrary threshold of at least three reads, of which at least one was required to be in the opposing orientation, detected a total of 23,262 variations, 17,031 of which were on chromosome 14 (73%, Table S1). Of these, 3,457 variants were found to not correspond to known SNPs, and were absent from the normal control individual (E). After initial exclusion of non-exonic and synonymous variants, 42 variants in 29 candidate genes remained. In 20 of these genes, a single variation was found in one individual, while two and three variations were found in six and two genes, respectively (Figure 3).

FLVCR2 was the only gene with variations identified in four out of seven individuals. In addition, careful examination of the *FLVCR2* locus in the proband of family II revealed a homozygous deletion of exons 2 to 10, as the absence of both nucleotide variations and reads over a 46.8 KB genomic region (Figure 4A). The deletion was confirmed to segregate in families I and II, and cloning of the breakpoints revealed the inclusion of the last two exons of the neighbouring *C14orf1* gene, with no repeated DNA sequences at the boundaries. It is noteworthy that this deletion was not detected by Affymetrix 250K SNP chip. Indeed, only one SNP was located in the non-deleted portion of intron 1. Direct sequencing of the 10 exons of *FLVCR2* (Supp. Table S1), identified mutations in two additional families (Table 1), such that mutant *FLVCR2* alleles were identified in each of

the 7 families studied (5 homozygotes and 2 compound heterozygotes; Table 1 and Figure 4B).

Reasons for false-negative results using HTS approaches are summarized in Table 1, and emphasize the need for complementary confirmation. In particular, in family IV, a second heterozygous mutation was found by direct resequencing, although it had an apparently homozygous mutation as indicated by the HTS analysis. In family III, the homozygous mutation found with Sanger sequencing had only been read 2 times in the HTS and had thus been excluded by the stringency of the filter. As a third example, the second heterozygous mutation in family VII had been read 4 times but was excluded for unidirectionality. Interestingly, in family VI, not known to be consanguineous, the identical nonsense mutation was found in the 3 affected sibs (homozygous in fetuses and heterozygous in parents), suggesting more distant consanguinity or a founder effect.

FLVCR2 is a member of the major facilitator superfamily (MFS) of transporter proteins, that shuttle small molecules in response to ion gradients (Pao, et al., 1991). Like other MFS members, FLVCR2 is predicted to contain 12 membrane-spanning segments and six extracellular loops. As shown in figure 5A, the 3 homozygous mutations are predicted to alter an amino-acid localized to one transmembrane domain (TM) : TM2 in family VI, TM8 in family III and TM10 in family V. In family IV, one of the 2 mutations alters an amino-acid predicted to be localized in TM8 and the other in the intracellular loop 5.

Amino acid sequence alignment for FLVCR2 from 10 different species showed that T430 and G412 have been conserved since our common ancestor with *C. elegans,* whereas R84 has been conserved in common with *D. melanogaster* (Figure 5B). T352R and L398V

alter residues less evolutionary conserved, especially L398V. However, those mutations are absent from both the dbSNP and the 1000 Genome database not yet integrated in dbSNP. While the L398V mutation was predicted to be benign by the Polyphen algorithm (http://genetics.bwh.harvard.edu/pph/), the T352R mutation as well as the other missense mutations identified in this study were predicted to be damaging to protein function. Thus, the pathogenicity of these two last mutations is likely but not totally proven. In total, eight different mutations including one nonsense mutation (homozygous in family VI), six missense mutations, and one homozygous deletion in two families (I and II) have been found in *FLVCR2*.

DISCUSSION

PGV is a very rare and lethal genetic condition. Since its first description, 42 cases from 26 families have been reported on the basis of histological criteria of PGV (Bessieres-Grattagliano, et al., 2009; Williams, et al., 2010). In the 16 fetuses of our series born to eight unrelated families, neuropathological analysis defined a diffuse form of encephaloclastic prolifrative vasculopathy (EPV), affecting the entire CNS and resulting in classical PGV with pterygia and a severe fetal akinesia deformation sequence in 14 cases. In contrast, two cases from the single family IV presented a more focal form of EPV, without spinal cord involvement and subsequent arthrogryposis/pterygia. Identification of *FLVCR2* mutations in this family suggests that the anteroposterior extent of CNS degeneration can be variable, and that PGV may be an extreme phenotype of a broader spectrum of proliferative vasculopathies. Stabilization of newly formed capillary

sprouts during angiogenesis requires interactions of endothelial cells with mural support cells, known as pericytes. The regionally restricted distribution of PGV in family IV might be linked to the embryonic lineage of the telencephalic pericytes, of a distinct neural crest cell origin from those of the spinal cord (Etchevers, et al., 2001). Interestingly, immunostaining for alpha smooth muscle actin (aSMA, a marker for mature pericytes) in fetal PGV brains was drastically reduced in the PGV within the CNS while normal aSMA expression was found in the leptomeninges (Figure 1I). Further studies should elucidate whether this observed effect on pericytes is the primary cause or an effect of this disease.

Recently, *FLVCR2* mutations were also reported in 5 families with Fowler syndrome (Meyer, et al., 2010), with the same homozygous Thr430Arg mutation in three families, and 2 compound heterozygous cases. Interestingely, Thr430Arg is associated with both forms of the disease, namely with or without spinal cord involvement, suggesting no genotype phenotype correlations. It is noteworthy that the mutation concerned the same codon (Thr430) as in our family IV, the only one of our series without spinal cord involvement. More recently, Lalonde et al. also reported four *FLVCR2* compound mutations in 2 FS families with spinal cord involvement ((Lalonde, et al., 2010). Interestingly, the only missense mutation predicted to be "benign" in our study (L398V) was identified by two distinct approaches in a common case reported by both Lalonde et al. and Meyer et al., adding to the likely pathogenicity of this variation. To sum up, 15 different *FLVCR2* mutations (including those described in our study) have now been reported in 13 cases: one large deletion, two nonsense mutations, one splice site mutation, one insertion/deletion change and 10 missense variations.

The *FLVCR2* gene encodes a transmembrane protein that belongs to the major facilitator superfamily (MFS) of secondary carriers that transport small solutes such as calcium (Pao, et al., 1991). It is closely related in both sequence and topology to the better-known FLVCR1, sharing 60% amino acid identity (Lipovich, et al., 2002). FLVCR1 has been identified as the receptor for a feline leukemia virus (FeLV-C), and like FLVCR2 and other MFS members, is predicted to contain 12 membrane-spanning segments and six extracellular loops. A single mutation in the sixth extracellular loop is sufficient to confer FeLV-C receptor activity on FLVCR2, which does not otherwise bind the native virus (Brown, et al., 2006). However, FLVCR2 functions as a receptor for the FeLV-C variant FY981 (Shalev, et al., 2009). FLVCR1 is found only in hematopoietic tissues, the pancreas and kidney (Tailor, et al., 1999), but rodent Flvcr2 is widely expressed during embryonic development, in particular within the CNS and in the vessels of the maturing retina, and human FLVCR2, within the fetal pituitary (Brasier, et al., 2004). FLVCR1 has been shown to function as a heme exporter, essential for erythropoiesis (Quigley, et al., 2004). Interestingly, the five glutamate residues in the C-terminal putative coiled-coiled domain of FLVCR2, not present in FLVCR1, may serve an analogous function to the same ferric ion-binding glutamate sequence in glycine-extended gastrin, by stimulating cell proliferation (He, et al., 2004). Based on the cell types in which it is expressed and MFS transport of chelated complexes of divalent metal ions, the FLVCR2 transporter was postulated to be a gatekeeper for the controlled entry of calcium into target cell types (Brasier, et al., 2004). Calcium signalling is involved in virtually all cellular processes and its homeostasis is tightly regulated. Angiogenic factors such as VEGF-A and FGF2 induce a transient increase of endothelial cell intracellular calcium concentrations , which

acts as a second messenger to induce proliferation, among other effects (Tomatis, et al., 2007). Blood vessels are susceptible to responding to angiogenic signals and undergoing calcification when their pericytic coverage has been disrupted (Collett and Canfield, 2005), both of which signs we have observed in PGV patient brain sections.

HTS of the entire exome has been used so far to identify disease-causing genes in the rare Miller and Bartter syndromes, respectively (Choi, et al., 2009; Ng, et al., 2010). Recently, targeted exon-specific sequencing within a restricted 40 MB linkage interval allowed the identification of an additional gene for Familial Exudative Vitreoretinopathy (Nikopoulos, et al., 2010). Our study underlines the use of HTS for the coverage of an entire linkage interval with no compelling candidate genes and no justification for the exclusion of non-coding regions. Our nested analysis approach led rapidly to the identification of a disease-causing gene. While it further demonstrates the power of this new technology, it also highlights other potential risks of missing mutations during data analyses. The number of patients, diagnostic accuracy and genetic homogeneity allowed us to compensate for low capture efficiency due to suboptimal DNA quality, and in the future, as the technology develops, furthering the depth of coverage should ensure a better distinction of background from true mutations. Finally, identification of the gene for Fowler syndrome will permit accurate genetic counselling for PGV and prenatal diagnosis, in particular for the late-onset forms of the disease without spinal cord involvement.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

ACKNOWLEDGEMENTS

We are grateful to families and to the French Society of Fetal Pathology (SOFFOET) for participating in the study. We thank Chantal Esculpavit for technical help. This work was funded by the GIS-Maladies Rares. ST is supported by grant NS039818 from the US National Institute of Health (NIH).

REFERENCES

Benko S, Fantes JA, Amiel J, Kleinjan DJ, Thomas S, Ramsay J, Jamshidi N, Essafi A, Heaney S, Gordon CT and others. 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat Genet 41(3):359-64.

Bessieres-Grattagliano B, Foliguet B, Devisme L, Loeuillet L, Marcorelles P, Bonniere M, Laquerriere A, Fallet-Bianco C, Martinovic J, Zrelli S and others. 2009. Refining the clinicopathological pattern of cerebral proliferative glomeruloid vasculopathy (Fowler syndrome): report of 16 fetal cases. Eur J Med Genet 52(6):386-92.

Brasier G, Tikellis C, Xuereb L, Craigie J, Casley D, Kovacs CS, Fudge NJ, Kalnins R, Cooper ME, Wookey PJ. 2004. Novel hexad repeats conserved in a putative transporter with restricted expression in cell types associated with growth, calcium exchange and homeostasis. Exp Cell Res 293(1):31-42.

Brown JK, Fung C, Tailor CS. 2006. Comprehensive mapping of receptor-functioning domains in feline leukemia virus subgroup C receptor FLVCR1. J Virol 80(4):1742-51.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S and others. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 106(45):19096-101.

Collett GD, Canfield AE. 2005. Angiogenesis and pericytes in the initiation of ectopic calcification. Circ Res 96(9):930-8.

Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16(22):10881-90.

Etchevers HC, Vincent C, Le Douarin NM, Couly GF. 2001. The cephalic neural crest provides pericytes and smooth muscle cells to all blood vessels of the face and forebrain. Development 128(7):1059-68.

Fowler M, Dow R, White TA, Greer CH. 1972. Congenital hydrocephalus-hydrencephaly in five siblings, with autopsy studies: a new disease. Dev Med Child Neurol 14(2):173-88.

He H, Shehan BP, Barnham KJ, Norton RS, Shulkes A, Baldwin GS. 2004. Biological activity and ferric ion binding of fragments of glycine-extended gastrin. Biochemistry 43(37):11853-61.

Kleinjan DA, van Heyningen V. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76(1):8-32.

Lalonde E, Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, Dechelotte P, Majewski J, Jabado N. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. Hum Mutat.

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12(14):1725-35.

Lipovich L, Hughes AL, King MC, Abkowitz JL, Quigley JG. 2002. Genomic structure and evolutionary context of the human feline leukemia virus subgroup C receptor (hFLVCR) gene: evidence for block duplications and de novo gene formation within duplicons of the hFLVCR locus. Gene 286(2):203-13.

Meyer E, Ricketts C, Morgan NV, Morris MR, Pasha S, Tee LJ, Rahman F, Bazin A, Bessieres B, Dechelotte P and others. Mutations in FLVCR2 are associated with proliferative vasculopathy and hydranencephaly-hydrocephaly syndrome (Fowler syndrome). Am J Hum Genet 86(3):471-8.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA and others. 2010 Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42(1):30-5.

Nikopoulos K, Gilissen C, Hoischen A, van Nouhuys CE, Boonstra FN, Blokland EA, Arts P, Wieskamp N, Strom TM, Ayuso C and others. 2010. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. Am J Hum Genet 86(2):240-7.

Pao GM, Wu LF, Johnson KD, Hofte H, Chrispeels MJ, Sweet G, Sandal NN, Saier MH, Jr. 1991. Evolution of the MIP family of integral membrane transport proteins. Mol Microbiol 5(1):33-7.

Quigley JG, Yang Z, Worthington MT, Phillips JD, Sabo KM, Sabath DE, Berg CL, Sassa S, Wood BL, Abkowitz JL. 2004. Identification of a human heme exporter that is essential for erythropoiesis. Cell 118(6):757-66.

Shalev Z, Duffy SP, Adema KW, Prasad R, Hussain N, Willett BJ, Tailor CS. 2009. Identification of a feline leukemia virus variant that can use THTR1, FLVCR1, and FLVCR2 for infection. J Virol 83(13):6706-16.

Tailor CS, Willett BJ, Kabat D. 1999. A putative cell surface receptor for anemia-inducing feline leukemia virus subgroup C is a member of a transporter superfamily. J Virol 73(8):6500-5.

Tomatis C, Fiorio Pla A, Munaron L. 2007. Cytosolic calcium microdomains by arachidonic acid and nitric oxide in endothelial cells. Cell Calcium 41(3):261-9.

Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE. 2008. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. Hum Mutat 29(1):6-13.

Williams D, Patel C, Fallet-Bianco C, Kalyanasundaram K, Yacoubi M, Dechelotte P, Scott R, Bazin A, Bessieres B, Marton T and others. 2010. Fowler syndrome-a clinical, radiological, and pathological study of 14 cases. Am J Med Genet A 152A(1):153-60.

LEGENDS TO TABLE AND FIGURES

Figure 1: Marker analysis in Fowler syndrome fetal brain

(A) Cortical plate of Fowler syndrome (FS) fetal brain (family IV) showing abnormal perforating vessels. Note the characteristic thickened vessels (asterisks), ending in glomeruloid formations (arrowheads), often devoid of recognizable lumina. CD34 capillary staining in (B) on a brain from a control, stage matched fetus and (C) from a FS fetus (family I) (C). VEGF immunostaining arround (D) a brain parenchymal capillary from a control fetus in which it is essentially absent, and (E) from a FS fetus in which it appears markedly increased. (F) GFAP astroglial immunostaining on a FS fetal brain. Alpha SMA immunostaining of pericytes on (G) a brain section from a control fetus versus (H and I) from two FS fetuses.

Figure 2 : Pedigree and linkage analysis results

Pedigrees of families included in this study. Arrows indicate individuals for whom DNA was available, and arrowheads indicate the samples sequenced by HTS. Homozygosity or linkage was analysed by microsatellite markers analysis and confirmed a founder effect by haplotype identity in 2 Turkish families (I and II) that were later discovered to carry the same *FLVCR2* exon 2 to 10 deletion.

Figure 3 : Summary of HTS data analysis

This diagram illustrates the flowchart of HTS data analysis. After elimination of variants found outside of the mapping region (27% of total variants) and those corresponding to known SNPs (29% of on-target variants) or shared with the control individual E (50% of on-target variants),

HTS identified 54 variants in coding sequences, eight of which were synonymous. The remaining 46 variants were located in 29 candidate genes, 20 of which were excluded because only one variant was identified. Finally, only one gene, *FLVCR2* presented 4 variants.

Figure 4 : *FLVCR2* deletion and mutations

(A) Genome browser view centered on the *FLVCR2* locus (ENSG00000119686) showing all variations (red dots) and reads coverage (light blue) in individuals A (fetus, family II) and B (fetus, family V). Note the absence of variations and reads in individual A, suggesting a homozygous deletion of exons 2 to 10, as well as the 2 final exons of the adjacent c14orf1 transcript (ENSG00000133935). (B) Chromatograms of *FLVCR2* homozygous (upper panel) and compound heterozygous mutations (lower panel).

Figure 5: Localisation of mutations in *FLVCR2* and conservation of mutated FLVCR2 amino acids.

(A) Localization of mutations on a secondary structure prediction of the FLVCR2 transporter. The three homozygous mutations are predicted to alter an amino acid localized in one of the 12 transmenbrane (TM) domains: p.Y134X is located in TM2, p.L359P in TM8, and p.G412R in TM10. Compound heterozygous mutations in family VI alter amino acids at the N-terminal cytoplasmic end and in the extracellular loop 5 (blue asterisk). Compound heterozygous mutations in family IV alter an amino acid predicted to be localized in TM8 and in the intracellular loop 5 (green asterisk).

(B) Alignment and conservation of mutated FLVCR2 amino acids. Sequences for FLVCR2 from 10 different species have been aligned using the Multialin tool ("Multiple sequence alignment

with hierarchical clustering" (Corpet, 1988). Highly conserved amino acids are represented in red, moderately conserved amino acids are in blue and non-conserved ones are in black. Mutated amino acids are boxed.

Table 1 : Analysis of variations by individual and *FLVCR2* variations identified by high-throughput or capillary sequencing.

E is a healthy brother in family V not carrying the disease allele by haplotyping, and taken as healthy control. Mutation numbering is based on cDNA sequence with a 'c.' symbol before the number, where +1 corresponds to the A of ATG translation codon (codon 1) of the cDNA reference sequences (NM_017791). Mutation names were checked by the Mutalyzer program (Wildeman, et al., 2008).

Supplementary Table S1: Analysis of total number of variations detected by high-throughput sequencing

Supplementary Table S2: Primers and PCR conditions

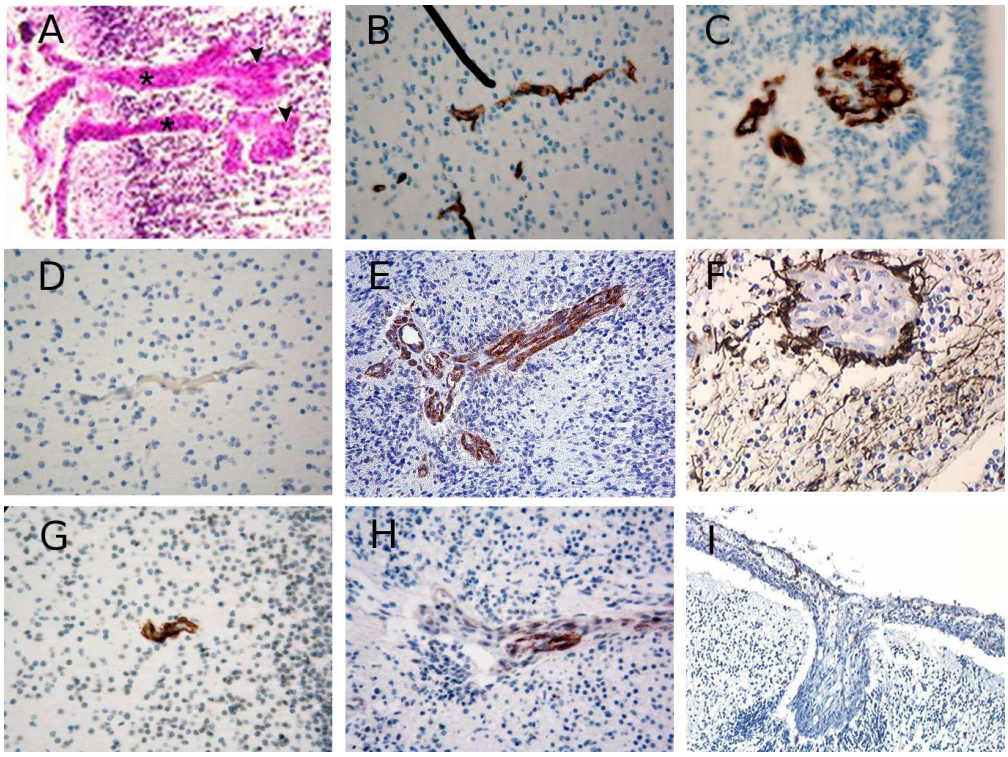


Figure 1
115x86mm (300 x 300 DPI)

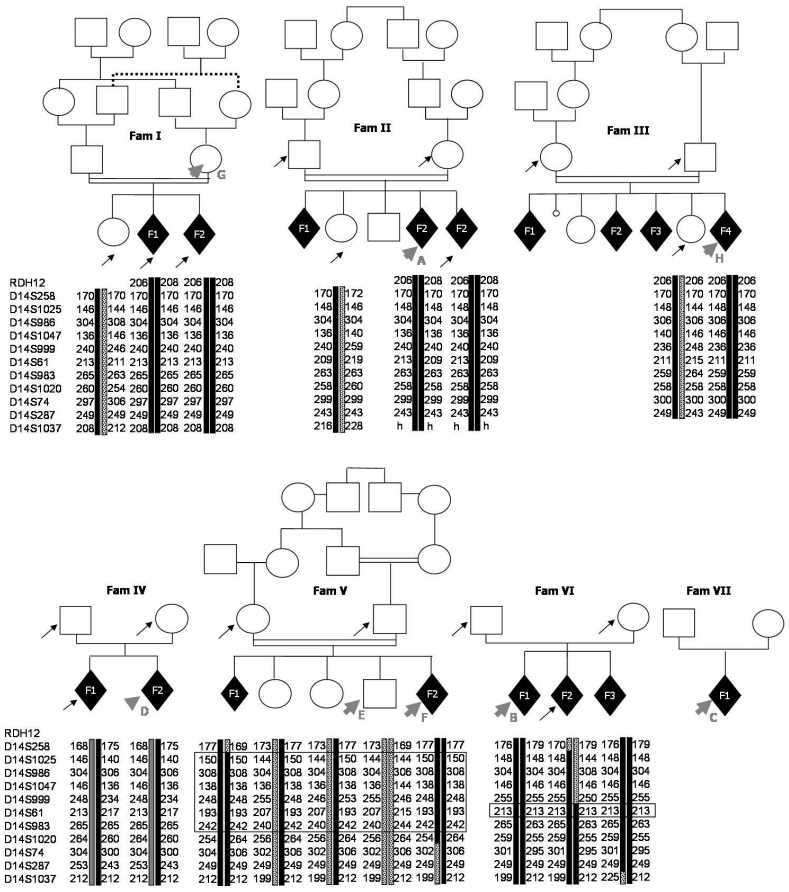


Figure 2
160x160mm (300 x 300 DPI)

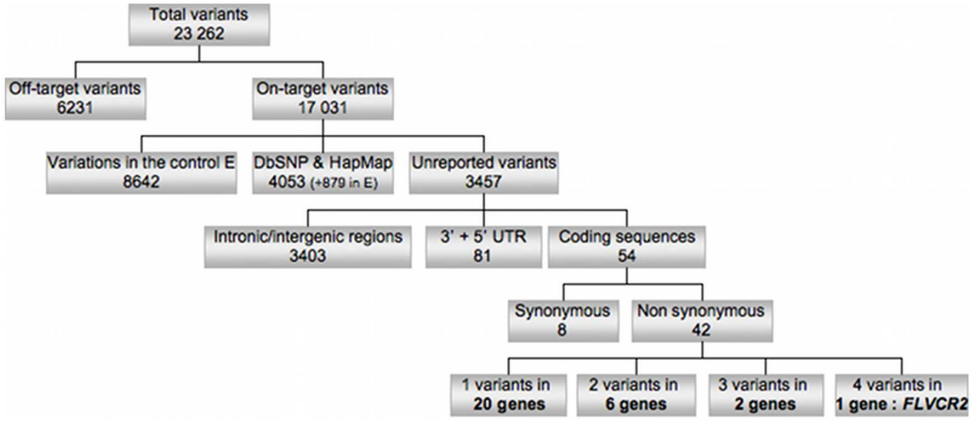


Figure 3
154x70mm (300 x 300 DPI)

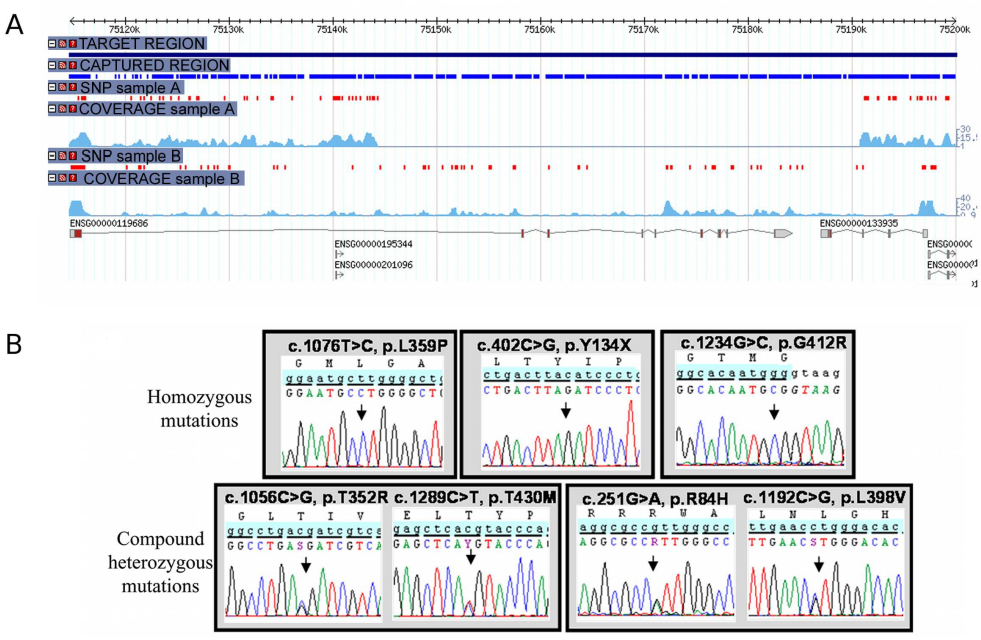


Figure 4
199x128mm (300 x 300 DPI)

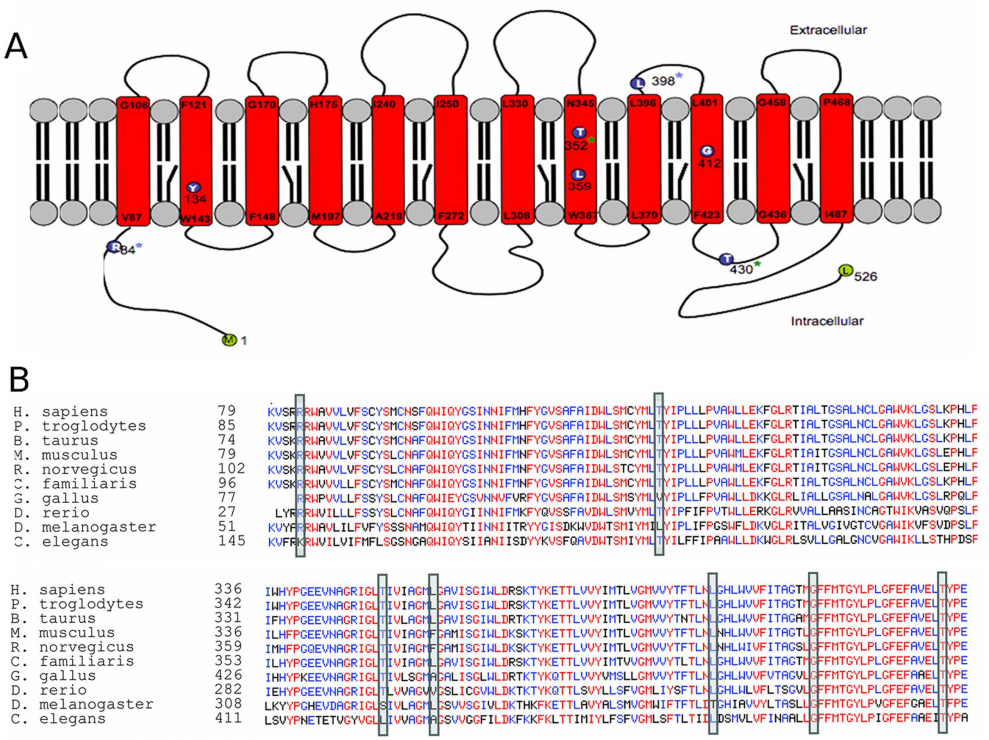


Figure 5
199x149mm (300 x 300 DPI)

	Individual	A (Fam II)	B (Fam VI)	C (Fam VII)	D (Fam IV)	F (Fam V)	G (Fam I)	H (Fam III)	*Total*
	Origin	Turkish	French	French	French	Maroccan	Turkish	French	
	Coverage	8,8X	4X	8,6X	2,3X	11,8X	11,6X	6,6X	7X
Number of Variations (total)	All	2852	1804	2639	823	3067	3841	2005	17 031
	Variations in E removed	1379	790	1154	282	1527	2075	1182	
	SNP removed	565	380	608	112	695	872	821	
	Variations in E and SNPs removed	546	300	465	80	569	750	747	
Number of variations on mRNA	Total	100	74	105	41	87	139	58	
	Variations in E and SNP removed	23	14	17	6	20	26	29	
Number of variations on CDS	Total	41	22	44	13	42	60	25	
	Variations in E and SNP removed	8	2	4	2	11	12	15	
	non synonymous	22	8	23	9	23	28	13	
	non synonymous and SNP removed	8	2	4	2	9	8	9	*42*
	Next generation sequencing	Del Ex 2-10 hmz	c.402C>G, p.Tyr134Stop hmz	c.251G>A, p.Arg84His htz	c.1056C>G, p.Thr352Arg hmz	c.1234G>C, p.Gly412Arg hmz	(mother)	-	
FLVCR2 variations	Capillary sequencing	Del Ex 2-10 hmz	c.402C>G, p.Tyr134Stop hmz	c.251G>A, p.Arg84His htz c.1192C>G, p.Leu398Val htz	c.1056C>G, p.Thr352Arg htz c.1289C>T, p.Thr430Met htz	c.1234G>C, p.Gly412Arg hmz	Del Ex 2-10 htz in mother	c.1076T>C, p.Leu359Pro hmz	
	Comparison and reason for discrepancy	confirmation	confirmation	Arg84His: confirmation Leu398Val: 7 reads, 4 with the mutation, but excluded for unidirectionality	T352R: 4 reads of only the mutated allele, T430M: no reads	confirmation	Deletion confirmed in foetuses (hmz), htz in parents	2 reads of only the mutated allele	

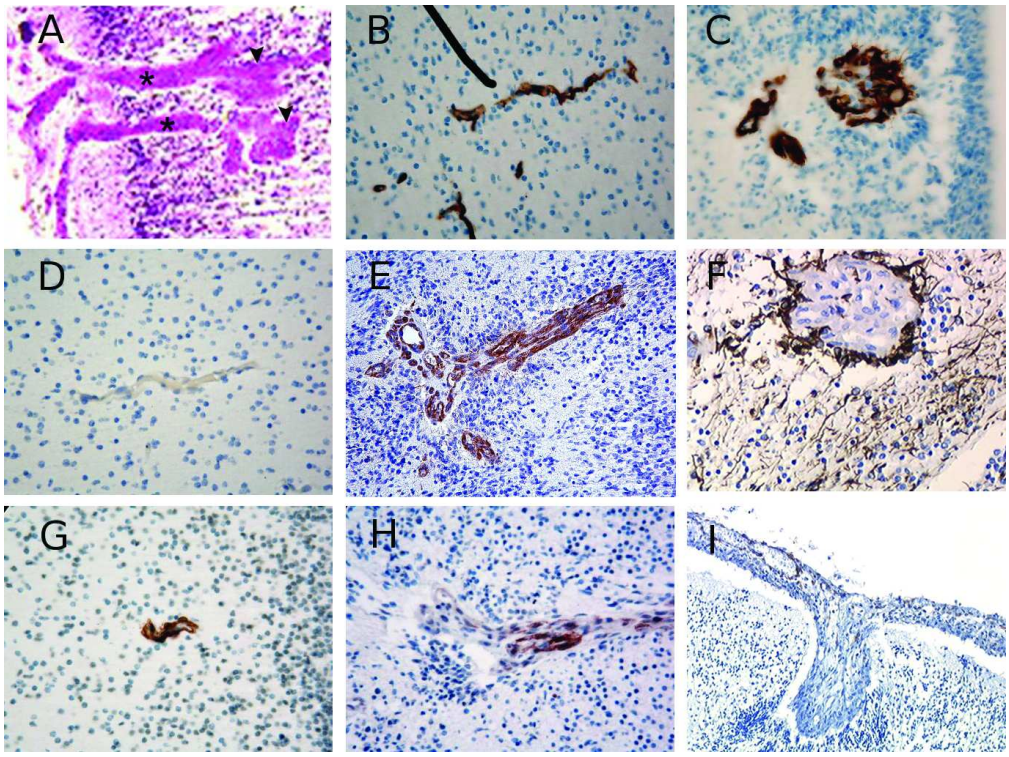
Table 1 : Analysis of variations by individual and *FLVCR2* variations identified by next generation or capillary sequencing. Mutation numbering is based on cDNA sequence with a 'c.' symbol before the number, where +1 corresponds to the A of ATG translation codon (codon 1) of the cDNA reference sequences (NM_017791). Mutation names were checked by the Mutalyzer program (Wildeman, et al., 2008).

Heterozygous and homozygous variations (20-100% of total reads)	**All variations**	**variations on chr 14**	**% on Chr 14**
A + B + C + D + F + G + H	23 262	17 031	73%
A + B + C + D + F + G + H **– E (control)**	12 942	8 389	65%
A + B + C + D + F + G + H **– E –HapMap**	10 177	5 763	57%
A + B + C + D + F + G + H **– E –dbSNP**	6 553	3 457	52%
A + B + C + D + F + G + H **– E –hapMap – dbSNP**	6 552	3 457	52%
A + B + C + D + F + G + H **– E –hapMap – dbSNP : mRNA (UTR et CDS)**	179	135	75%
A + B + C + D + F + G + H **– E –hapMap – dbSNP : CDS**	63	54	85%
A + B + C + D + F + G + H **– E –hapMap – dbSNP : CDS, non synonymous**	49	**42**	85%

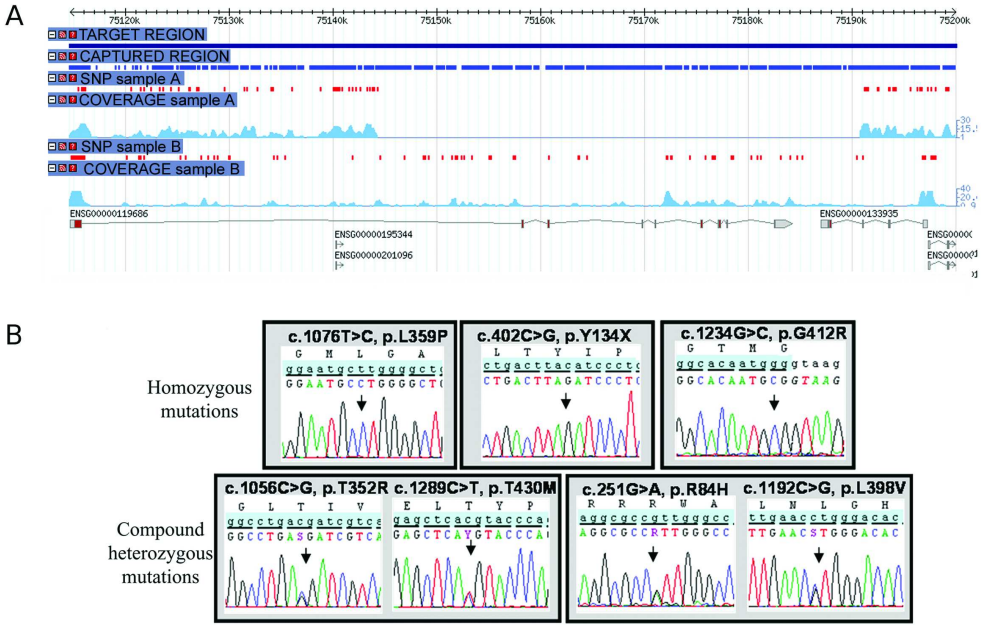
Supplementary Table 1
Analysis of Total Number of variations

Primers	Sequence	PCR size
FLVCR2-1AF	GCCTCTAGTCTCTGTTTCTTCTGG	527
FLVCR2-1AR	TCAGCATGTAGCACATGGAC	
FLVCR2-1BF	TGTGCAACTCCTTTCAGTGG	527
FLVCR2-1BR	CAATCACTGCCTGTCACACC	
FLVCR2-2F	TCTCTGGTGTTTTGAGGTGAGA	397
FLVCR2-2R	CATGGTATTTCAGGGCATGTT	
FLVCR2-3F	TTCACTCAGCCTCAAACAATG	
FLVCR2-3R	TAGCTGGGTCCTCTGGATTG	
FLVCR2-4F	TGTGTGGCTAAGGGAAGGTT	464
FLVCR2-4R	GGTTGAGATCTAGGGCCATCT	
FLVCR2-5F	TCTCCTAGGCCATCTTGTCC	363
FLVCR2-5R	CTTGGCCACTAGGATCTCCA	
FLVCR2-6F	GGCAACAGAGCAAGACACTG	382
FLVCR2-6R	TCAGTTAGAAGGCAGCAAAGG	
FLVCR2-7F	CCCAGATCATTAGAGGGCCTA	596
FLVCR2-8R	CCAACAAACCCTTCCATCTG	
FLVCR2-9F	CCTGTGACCCTTAGGAAATGA	292
FLVCR2-9R	TGCCATGTGTAAGGGATGAA	
FLVCR2-10F	TTTCTTGGCTCTCTGGGATG	486
FLVCR2-10R	TATTCTCTGCCACCCTGTCC	
Primers used for cloning the deletion breakpoints of families I and II		
FLVCR1-i1	CAGGATAAGCTCCATCATCCTTAC	
C14orf1-3Fex	CTCGGACCTTTGGGATCTG	

Supplementary Table 2
Primers and PCR conditions



115x86mm (300 x 300 DPI)



199x128mm (300 x 300 DPI)

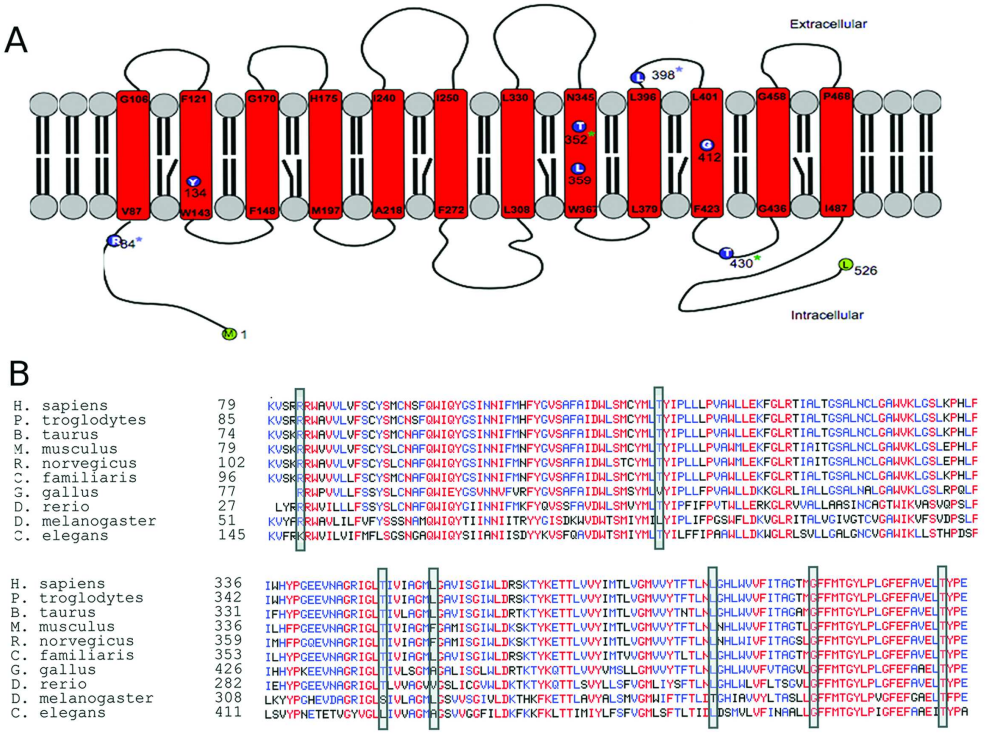


Figure 5
199x149mm (300 x 300 DPI)